



Evidence vs. Professional Judgment in Ranking “Power Few” Crime Targets: a Comparative Analysis

James Sutherland, et al. *[full author details at the end of the article]*

Published online: 14 May 2019
© The Author(s) 2019

Abstract

Research question How accurately can local police officers use professional judgement to identify the highest-crime street locations and offenders with the most crime and harm, in comparison to an evidence-based rank-ordering of all possible locations and names derived from police force records?

Data A face-to-face survey was conducted in groups with a purposive convenience sample of 123 operational police officers to ask their professional judgement for selecting the ten most crime-prone streets and suspected offenders in their command areas. Separate rankings by crime harm were also requested. Cambridgeshire Constabulary crime and confirmed suspect reports were analysed to create the same lists the officers were asked to provide.

Methods The study compared results of surveys of police officers asked to name the top 10 streets and offenders for volume and harm of crimes committed in each policing area to the top ten lists generated by comprehensive and systematic analysis of reported crimes.

Findings The top ten lists generated by officers were highly inaccurate compared to the lists produced by comprehensive analysis of crime and charging records. Officers surveyed were 91% inaccurate in naming the most prolific suspected offenders in their areas and 95% inaccurate in naming the most harmful suspected offenders. Officers were slightly less inaccurate in naming the streets in their areas with the highest frequency of crimes (77% incorrect) and the greatest severity of crimes (74% incorrect). Officers in urban areas ($N=42$) were substantially more accurate than officers working in semi-rural areas ($N=30$) in identifying streets with the highest crime frequency (Cohen's $d=0.9$; $p=.00$) and highest total harm (Cohen's $d=1.3$; $p=.00$), but urban officers still failed to name about two-thirds of the most harmful streets.

Conclusions Police officers can benefit from evidence-based targeting analysis to help them decide where their proactive and preventive work can be deployed with the greatest benefit.

Keywords Power few · Hot spots · Repeat offenders · Professional judgement · Intuition

Introduction

Police and criminologists increasingly agree that criminal events are heavily concentrated in a tiny minority of all possible locations, offenders and victims. The subject of this study is whether they can agree on *which* locations or offenders have the *highest* concentrations for purposes of police resource allocation. Because operational police usually rely on professional judgement, while empirical criminologists rely on systematic analysis of police records, there is a clear possibility that the different methods will yield different results. Thus, the question is not really about *who* is identifying the high-crime targets, but *how* the identification is done. The answer we report is that comprehensive statistical evidence identifies very different targets than selection based on professional judgement of officers working in the same local areas.

The concentration of crime into ‘hot spot’ locations has been well established through years of research in a range of countries and environments (Sherman et al. 1989; Weisburd 2015). It is also well established that some offenders are more prolific than others and responsible for disproportionate amounts of crime (Sherman 2007). Consequently, policing strategies that target hot spots (Braga et al. 2012) and the most prolific serious offenders (Martin and Sherman 1986) have become well-established policing tactics whose efficacy has been widely accepted.

Although less established, a growing body of research has also found concentrations of total *harm* from crime, as distinct from counting all crime events as if they were of equal seriousness. Based on the idea of a crime-harm index (Sherman 2007; Sherman et al. 2016), these studies make possible the identification of *harm-spots* (Weinborn et al. 2017) as well as hot spots. They also draw attention to *offenders* who cause the most detected harm from crime (Liggins 2017) and not simply those who commit the most detected criminal events. The dimension of crime harm is especially important for victims, less than 4% of whom may suffer 85% of all of the weight of crime reported to a police agency in a single year (Dudfield et al. 2017).

Whether such systematic evidence can supplement the professional judgement of police officers, however, is not an easy question to answer. Many journalists ask why data analysis is even necessary, since “police already know from their own experience where the hot spots are.” The need for data analysis has also been questioned by police officers who distrust outside experts and academic research. Given their implicit trust in the value of experiential learning, they prefer to use that experience in determining the targets of their proactive patrol work, sometimes irrespective of the views and directives of police managers—let alone the results of data analysis.

The legal and cultural power of police discretion to shape police operations makes it vitally important to address their scepticism about evidence-based targeting. Given the strategic value of identifying any “power few” list of priority targets, there should be great value in a comparison of the two methods of identifying such targets. As long as the choices of targets are based solely on professional judgement, there is potential for both disagreement and error, as a recent study in Northern Ireland has shown (Macbeth and Ariel 2017). Given police responsibility to exercise their discretion in where to patrol, it behoves police management to find ways that lead them to the most accurately identified hotspot locations of crime and of harm, as well as to identify the most prolific and harmful offenders.

Professional Judgement and Confidence in Decision Making In his summary of evidence on the accuracy of human decisions, Kahneman (2011) does suggest occasions on which professional expertise and intuition *can* be trusted to produce optimal decisions. The requirements he proposes for expert skill to be developed are (1) an operating environment that is regular enough as to be predictable combined with (2) an opportunity to understand this regularity through practice. Married to these requirements is (3) the necessity of timely feedback on activities to reinforce the learning—although in situations of heightened danger, the learning can be achieved on single instances. Kahneman posits the example of the experienced fire-fighter whose intuition as to when a building is about to collapse can be relied upon. The experience of front line officers in situations of conflict would be a good comparator to this example. Frequent exposure to aggression would allow officers to recognise intuitively the warning signs of imminent attack—building on the expertise that any individual naturally has in this regard through utilisation of ‘the gift of fear’ (De Becker 2000).

However, an intuitive understanding of stable patterns is unlikely to meet the three requirements for accurate decisions Kahneman (2011) proposed when the task is selecting from a large universe a small number of policing targets: hotspots of crime, vulnerable victims or the most prolific offenders. Just as experts at stock market investment rarely do better than chance, experts at picking out the needles in a haystack of crime cannot rely on a stable operating environment. Nor can practice make perfect if the environment keeps changing. Moreover, officers will not have direct experience of the majority of incidents of crime.

Evidence on Professional Judgement About Hot Spot Locations Several previous studies have compared professional judgement to systematic data analysis. One early test used a geographical information system to identify hotspots of a limited number of crime types in an area (Ratcliffe and McCullagh 2001), which was then contrasted with perceptions of officers on the location of hotspots through surveys and focus groups. The findings suggested great variance between police recognition of hotspots and systematic analysis of reports of criminal activity. The authors suggested that the sheer volume of crime was too great for officers to be able to process ‘objectively’ whilst at the same time, they were prone to bias caused by attending traumatic incidents—as suggested by Kahneman (2011).

In a study by Rengert and Pelfrey (1997), a comparison was drawn between the impressions of police cadets of relative safety of different neighbourhoods in Philadelphia with the objective reality of safety in those areas. The authors found that the cadet-officer perceptions were divergent from the objective reality of dangerous places, with safe areas perceived as dangerous and vice versa.

More recently, Macbeth (2015) compared the presence of hotspots of crime in Northern Ireland identified by computer analysis with the supposed hotspots identified using ‘waymarkers’ by officers based on their judgement and professional experience. More than 97% of the streets identified by officers as hotspots were false-positives: they were not in fact hotspots of either crime counts or crime harm (Macbeth and Ariel 2017). Conversely, 60% of streets which analysis revealed were hotspots were not included within the waymarkers—creating a more worrying problem of a large number of false negatives where the opportunity to reduce harm was missed.

Research Questions

The primary focus of this research concerns the accuracy of police officer intuitions in the identification of hotspots of crime and crime harm. The research questions therefore are

1. How accurate is police officer judgement at identification of hotspots of crime counts in their area compared to a systematic analysis of all reported crime over 3 years?
2. How accurate is police officer judgement in the identification of hotspots of crime harm in their area compared to a systematic analysis of all crime over 3 years?
3. How accurate is police officer judgement at identification of prolific suspected offenders in their area with respect to total crime suspected compared to a systematic analysis of all offences of each confirmed suspect across all confirmed suspects over 3 years?
4. How accurate is police officer judgement at identification of offenders in their area suspected to have caused the most harm, compared to a systematic analysis of all crime categories across all crimes by all confirmed suspects over 3 years?

The following sub-research questions were also examined to help explore the primary research question, namely:

1. To what extent are crime counts concentrated across space within Cambridgeshire?
2. To what extent are crime counts concentrated across space in urban, rural and semi-rural areas?
3. To what extent is crime harm concentrated across space within Cambridgeshire?
4. To what extent is crime harm concentrated across space in urban, rural and semi-rural areas?
5. What is the relationship between the years of experience of a police officer and the accuracy of their professional judgement on crime hotspots, harm hotspots and prolific and harmful offenders?
6. To what extent does the length of experience of a police officer of working in *a specific geographical area* affect the accuracy of their professional judgement in selecting on crime hotspots, harm hotspots and prolific and harmful suspected offenders?
7. To what extent does the confidence that Police Officers have in the accuracy of their intuitions correlate with the actual accuracy of their predictions?

Data

The Setting At the time of this study, Cambridgeshire Constabulary was divided into six command areas—Cambridge City, Peterborough, Fenland, Huntingdonshire, East Cambs and South Cambs. At the time of writing, the first author was policing commander for South Cambs which allowed immediate access to both participants and data and therefore made an ideal piloting location. Furthermore, South Cambs is unique amongst the six command areas in being purely rural in nature—consisting of

105 parishes/villages and no towns. It was therefore used solely for a pre-test, the purpose of which was to refine the measurement procedures for the main study undertaken in the remaining five, more urban or town-centred, command areas.

Pre-testing For the pre-test only, 1 year of crime data was extracted from the Constabulary data-warehouse and reproduced in an excel-spreadsheet. Analysis of these data showed that over a 1-year period,

- just under 50% of all streets in the area were completely crime free.
- 3% of streets suffered 31% of crime.
- a ‘power-few curve’ distribution, as predicted by the crime-hotspots literature, was revealed for this rural area.

Having identified and rank-ordered the hottest streets for crime, the author then began a series of workshops with front line officers (both police officers and Police Community Service Officers). The participants were gathered into impromptu small-group workshops and provided pens and paper. They were then asked, individually and without conferring or referring to computer systems, to rank order the ten worst streets for crime in the area (where worst was explained as having the most crime, rather than any consideration of severity of crime or presence of offenders). During these workshops, which usually consisted of 6–10 officers, the officers appeared visibly ‘pained’ in trying to rank streets—with actual ‘head scratching’ and furrowed brows in effect. Officers had a near irresistible urge to confer or consult maps, despite being requested not to do so. This suggested to the first author that any future design would have to be done in a small group or 1–1 setting, rather than by remote survey, in order to prevent the development of a group/team consensus. When the 25 responses across four workshops were compared to the systematic tabulation of crimes by streets, the accuracy of professional judgement answers was extremely low. On average, officers named only two of the ten locations on the list of top-ranked streets in the command area.

At the conclusion of the pre-testing, the first author instigated a reform to local policing practice whereby PCSO patrols and action plans were re-directed to focus on the identified hot streets. This was widely publicised and promoted locally, leading to a far greater level of front-line officer awareness of the location of hot spot streets. Given this history, however, the area of South Cambs was excluded from the final study, due to the contamination of these testing effects. Yet, there was no evidence of spillover of these testing effects into the other five areas. No other command area within the constabulary carried out a similar analysis or adopted a similar approach, making the remaining five command areas suitable areas for further study.

The main study, after excluding South Cambs, was conducted in the five other command areas: Cambridge City, Peterborough, Fenland, Huntingdonshire and East Cambs. The main study was conducted in two *stages*, in which the first stage had two tranches of data collection. Tranche 1 of stage 1 was the identification of high crime and crime-harm hotspots. Tranche 2 of stage 1 compiled the identities of confirmed suspects as the offenders in any of those crimes, by area.

The stage 2 of data collection consisted of in-person surveys of officers working in each area to seek their identification of the same “top-ten” lists of

policing targets in their areas, based on each officers' professional judgement and experience in those areas.

Stage 1: Recorded Crime Analysis of Hotspots Suspected Offenders of Highest Crime and Harm

The data set was requested and received in two tranches. Tranche 1 contained the spatial evidence on crime and harm by locations, described in the following variables: date of offence, location of offence—both street and town/village—Home Office [national government crime categories] code and sub-code, offence description, which of the five relevant command areas and a unique crime reference number (automatically generated by the crime-file program).

Tranche 2 contained the same information but only included crimes where a named offender had been identified on the crime; this tranche is discussed below.

Tranche 1, Stage 1: Analysing Places

The following steps were performed with tranche 1 to achieve a rank ordered list of high crime count and high crime-harm locations. (1) The street data were segmented between the six different areas. (2) The data for crimes on *all streets that had at least one crime in the 3-year period* were ordered alphabetically by street name, and all locations that were major multi-lane road routes were identified and removed from the data set (for example, the road 'A14' which is a major, if not official, UK Motorway). Officers were informed of this fact in their instructions in part 2, detailed below. (3) The number of offences that had taken place within one street was aggregated and the streets were rank ordered according to the highest to the lowest levels of crime. This procedure was carried out for all six command areas. The issue of duplicate road names ("The High Street") was overcome by combining street-name with the town/village into one single category. At the end of this process, the first author had achieved a data set of every street in the county (which suffered at least one crime in the 3-year period) rank ordered from 'hottest' to 'coldest' according to crime count.

In order to weight each crime by the severity score of the Cambridge Crime Harm Index (CCHI), a similar method was used to that described in the last paragraph about ranking streets by counting all crimes as having equal weight. Utilising the CCHI spreadsheet made available on the Cambridge Institute of Criminology web-site (<https://www.crim.cam.ac.uk/Research/research-tools/cambridge-crime-harm-index/view>), the CHI value of each recorded crime was inserted into the original Cambridgeshire Police data set by replacing the Home-Office code/sub code with the corresponding CHI value. This process was not unproblematic (see Sutherland 2017).

Tranche 2, Stage 1: Analysing Confirmed Suspects

In tranche 2 of stage 1, the ranking process described above was repeated in a similar fashion to analyse the concentration of crime counts and harm severity by any and all confirmed suspects in the crime records (see Sutherland 2017). This tranche 2 analysis first tackled the issue of duplicates of named suspects. Then, the crime counts for each

suspect were summed within each individual suspect's row, so that all suspects could then be ranked in order from the most to the least prolific by the number of offences in which they were named in the offence reports as confirmed suspects.

Next, the CCHI value of each suspected offence by each suspect was entered into the data set, so that an aggregated total CCHI score for each suspect was calculated. All suspected offenders were then rank ordered according to how much total CCHI weight was associated with all of the offences for which they were confirmed as suspects (measured by recommended days of imprisonment for first offenders for each offence; see Sherman et al. 2016).

At the conclusion of this process, the first author had compiled a data set compiling every street on which one or more offence took place over 3 years, with the identifiers of every suspected offender (for every one of the offences with any confirmed suspects recorded) in the county over a 3-year period, so that four rank-ordered lists could be generated for 3-year (1) crime frequency by street, (2) CCHI by street, (3) crime count by confirmed suspect and (4) CCHI by confirmed crime suspect, all of which could be aggregated or subdivided by police command area.

Stage 2: Officer Professional Judgement for Identifying Places and Offenders of High Crime and Harm

In stage 2 of data collection, a completely separate, second data set was assembled to record responses to questions asked of individual police officers working in each area. These data were assembled by the first author conducting in-person surveys with 126 officers spread across five areas. The sampling frame for this stage of the research was the complete list of frontline uniformed 'response' officers and Police Community Service Officers (PCSOs) within Cambridgeshire Constabulary (outside of the excluded pre-test area, South Cambs). The strategy was to capture a non-probability stratified convenience sample that balanced numbers of officers from both urban and semi-urban police command areas. 'Urban' was defined as police command areas that centre almost entirely on urban centres of population, i.e. Peterborough and Cambridge. "Semi-rural" was defined as those police command areas where officers cover small-to-medium sized market towns and large numbers of villages, i.e. East Cambs, Fenland and Huntingdonshire.

The rationale for adopting a non-probability sampling strategy was as follows. The first author concluded from the pre-test that the best data collection strategy would be to conduct the survey in small workshop groups, face-to-face. However, at this point, practical considerations were faced: recent increases in demand meant that all non-essential training, meetings and secondments were cancelled during the research period. In such an environment, requiring a selected list of officers who met the requirements for a probability sample to attend face-to-face workshops was impracticable and unlikely to receive organisational support.

The adopted methodology was to 'piggy-back' on existing shift-briefings. Held daily, these briefings were an opportunity to survey large numbers of officers simultaneously without removing them from frontline duties or arranging to see particular officers at specified times. The only principle in attending various briefings was to obtain roughly equal numbers of responses from urban and semi-rural areas.

Sampling PCSOs presented greater practical difficulties as they did not routinely attend shift briefings. The author therefore adopted an ‘accidental sampling’ method (Hagan 2006) where PCSOs were sought out station by station based on their availability and surveyed in small group or one-on-one settings.

Following a sampling-size heuristic for the minimum number of cases in each category (Field 2013), the thesis sought to achieve a minimum of 30 officers in each category.

Given the relatively small number of PCSOs within the constabulary and the practical difficulties in assembling large numbers of PCSOs together due to the lack of team briefings and disparate locations, no distinction was made between urban and rural PCSOs. Due to these difficulties, only 14 PCSOs were successfully sampled during the research, a limitation that means some caution must be given to findings relating specifically to PCSOs.

The sample size for the workshops concerning prolific and harmful offenders was determined differently. The research questions in relation to offenders do not distinguish between urban and semi-rural areas and so the total population number of officers was larger. Again, following Field (2013), a minimum sample size of 30 was sought; the workshops yielded 40 participants. In total, for both places and offenders therefore, the author carried out workshops with a total of 123 officers and PCSOs across the force (no officer participated in more than one workshop).

Survey Procedure The first author, as a senior police leader, personally conducted all of the surveys with workshops and individuals. The format of the workshops was as follows: (1) the participants were given a brief overview of the research and its aims; (2) participants were given assurances that their responses would be entirely anonymous and would not be used as an individual assessment of their abilities or knowledge; (3) the participants were issued with hard copy answer sheets requesting the following information on their police role (response officer/PCSO), length of service (years/months) and how long the participant had been based in *this role in this area*; (4) the respondents were asked to fill out the answer sheet in response to two questions:

Scored out of 100% (where 100% is total confidence and 0% is no confidence at all)
how confident are you that you can identify the top ten crime hotspots in this area?

The second question was

Scored out of 100% (where 100% is total confidence and 0% is no confidence at all)
how confident are you that you can identify the top ten crime-harm hotspots in this area.

These questions on confidence were intended to obtain a subjective/intuitive response from the participants on their level of confidence in their knowledge of crime-hotspots.

The remainder of the sheet contained two numbered columns, containing black spaces next to the numbers 1–10. The participants were first asked to rank the top ten streets in the left hand column according to the total number of crimes that took place in that street. Participants were advised that if it aided them to rank targets from the worst

to the tenth worst, then they should do so but that the study was not concerned about the relative positioning within the top ten (in other words, the top ten streets could be in any order). A copy of the question sheet used within the workshops can be found in Sutherland (2017, Annex A).

Following this, the participants were asked to rank order the top ten streets in terms of crime-harm. The same advice on relative placing within the top ten was given in addition to a short explanation of the meaning of the term ‘harm’: participants were advised that harm meant the seriousness of a crime with the example given ‘a burglary is more serious than a cycle theft’. Participants appeared to understand this instruction without difficulty.

The workshops lasted approximately 15 min. Participation was 100%; no individual requested not to take part. The author was able to prevent reference to materials. This was important as the workshops were set up to mimic, as closely as possible, the intuitive professional judgement that response officers routinely use when out on patrol.

When scoring the answer sheets, each participant was given a mark out of ten for both crime count and crime harm lists, with one mark being awarded for a street correctly identified as being within the top ten, irrespective of the relative positioning within the top ten giving each participant an ‘intuitive accuracy’ score on a possible range of between 1 and 10. This method of measurement was not unproblematic, nor are the ethical issues it entailed; both are discussed at length in Sutherland (2017).

Findings

Hotspots of Crime Counts

Following the methodology described above, the number of crimes over a 3-year period were summed for each street, and all streets that had one or more crimes were rank ordered from highest to lowest in number of crimes. The number of streets with crimes, and the percentages of all crimes on each street with crimes were then calculated and tabulated as cumulative percentages from the highest count street to the lowest (cf. Sherman et al. 1989: Table 1, except for the limitation in this study to streets with one or more crimes). On a county wide basis, 5% of all the streets over 3 years that had one or more crimes accounted for 51.4% of the crime events.

Note that while this result is very close to Weisburd’s (2015) “law of crime concentration,” it actually *understates* the degree of concentration of crime across units because it omits streets that had no crime, while Weisburd’s review included street segments that had no crime at all (see also Sherman et al. 1989: Table 1). On the other hand, the bias towards under-estimating concentration in the present study is balanced by the fact that Cambridgeshire streets may occupy far more land mass than the street “segments” that form the unit of analysis in Weisburd’s review, thus providing more space within which crimes can be committed. Nonetheless, the main aim of the present study is not to estimate concentration, but rather to estimate whether police can recall the streets to which their police agency overall is called to record crime most often. And for that purpose, the study contains 100% of the relevant streets.

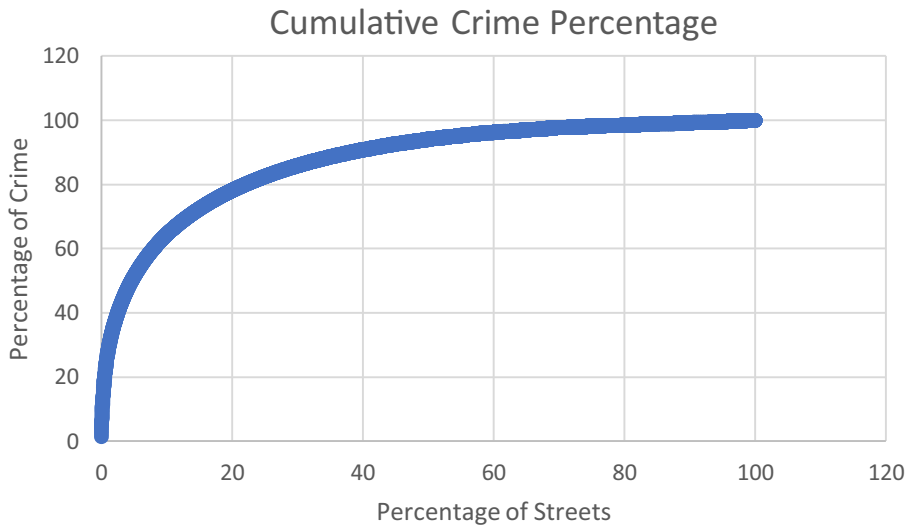


Fig. 1 Crime concentration across all Cambridgeshire

When the concentration of all crime events across all streets with any crime is displayed as a graph, the cumulative street and crime percentages reveal the existence of a ‘power curve’ in which the ‘power few’ are located at the far left hand side (Fig. 1):

Further analysis revealed that a similar pattern of crime concentration was found in all six command areas within Cambridgeshire, including urban, semi-rural and rural (see Figs. 2, 3, 4, 5, 6, and 7). Within the urban areas of Cambridge and Peterborough, the top 1% of streets suffered 25% and 27%, respectively. Within the semi-rural areas of Huntingdon, Fenland and East Cambridgeshire, the top 1% of streets suffered 20%, 20.2% and 15.5% of crime, respectively. Within the only purely rural area, South Cambridgeshire, the top 1% of streets suffered 17% of all crime.

Again, when graphed, the predicted ‘power-few’ curve can be found in all policing areas (see Fig. 8):

When visualised on the same graph, the recurring pattern of crime concentration in all areas of Cambridgeshire is even clearer.

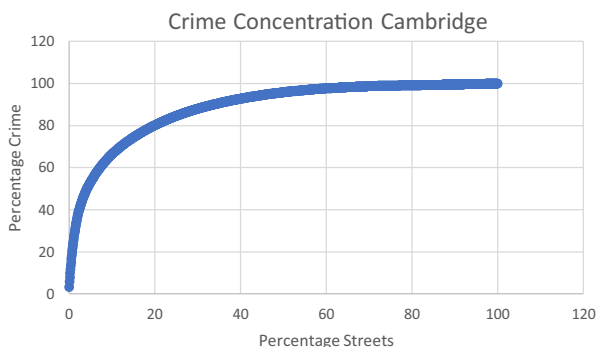


Fig. 2 Crime concentration in Cambridge City

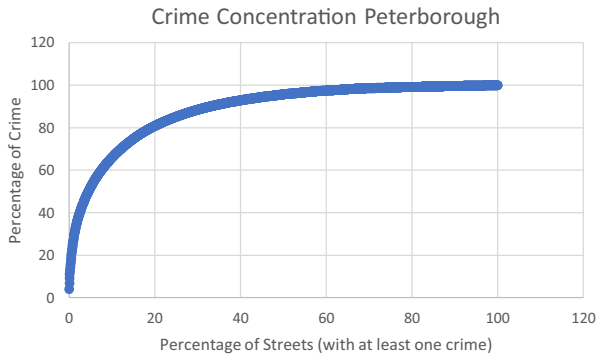


Fig. 3 Crime concentration in Peterborough

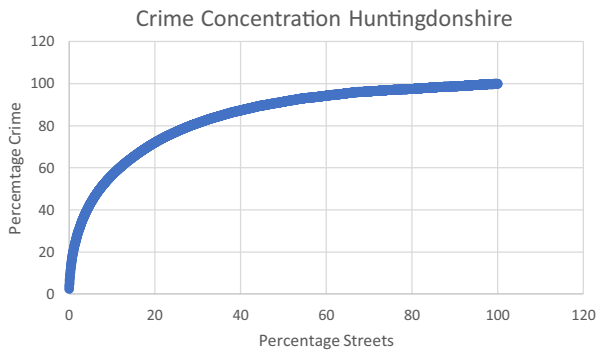


Fig. 4 Crime concentration in Huntingdonshire

Hot Spots of Total Crime Harm

Following the same methodology as with crime counts, the next analysis aggregated and summed the total Cambridge Crime Harm Index score (Sherman et al. 2016) across all of the crimes reported for each street. The total days of recommended imprisonment was used to rank streets from highest to lowest, with the cumulative distribution of

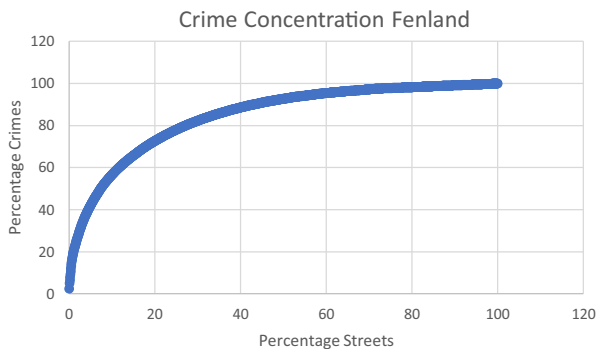


Fig. 5 Crime concentration in Fenland

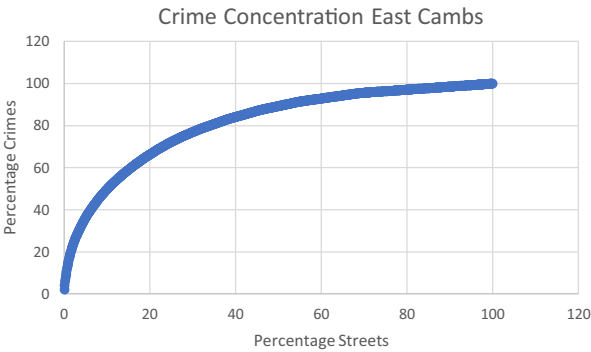


Fig. 6 Crime concentration in East Cambridgeshire

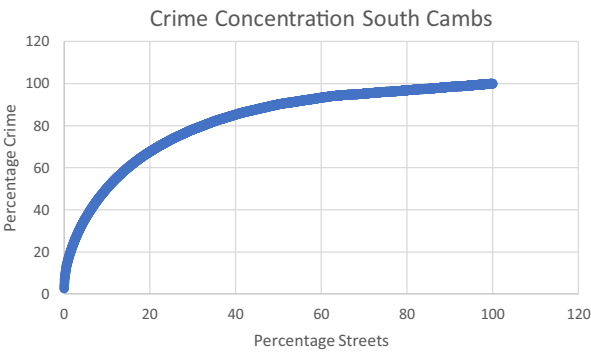


Fig. 7 Crime concentration in South Cambridgeshire

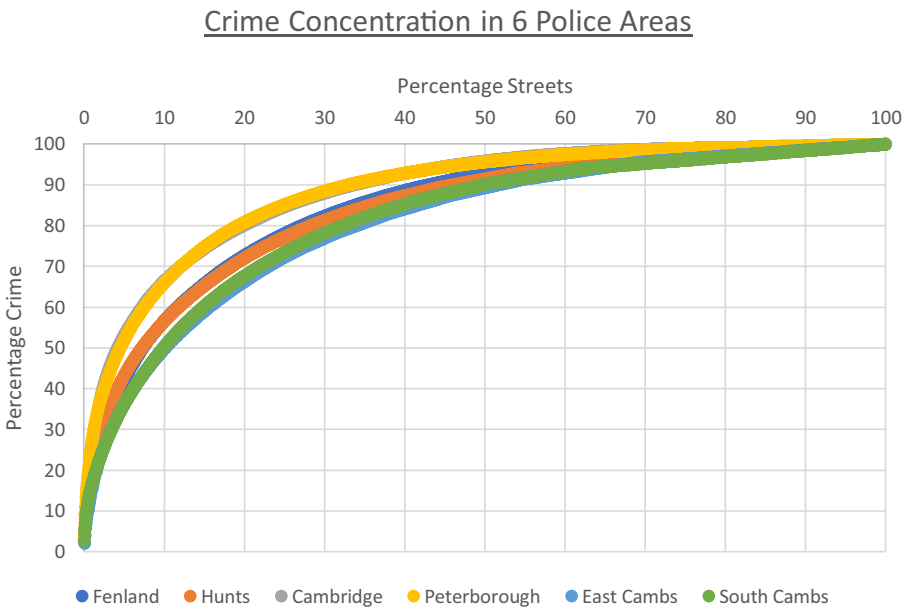


Fig. 8 Crime concentration in six police areas

harm across the streets displayed in Fig. 9 below. Across all of Cambridgeshire, 5% of the streets generated 53% of total Cambridge CHI crime harm.

Similar concentrations of crime harm were found in each of police areas within Cambridgeshire, in urban, rural and semi-rural areas (Sutherland 2017: Table 6).

Top-Ranked Prolific Offenders

Using the same research methodology as with streets, the analysis of detected offenders aggregated the count of offences on which offenders were confirmed as suspects over a 3-year period. We then calculated and aggregated the CHI score for those offences, giving each individual offender both a total count of crimes committed and a total weight of recommended imprisonment for the Cambridge CHI score. The research identified a total of 21,151 unique confirmed suspects, of whom the top 5% (“power few”) accounted for 27% of the criminal events with named suspects, while (a somewhat different) 5% of confirmed suspects accounted for 66% of the total CHI score for the harms caused by crime (Sherman et al. 2016). Figures 10 and 11 show the greater concentration of harm within the power few than was found for the concentration of detections, in which all detections are given equal weight.

What can be observed from these data is that while both crime counts and total harm are concentrated in a minority of confirmed suspects, those concentrations are far more pronounced in relation to harm than to counts of crime. The analysis found that 95% of all harm was caused by just 12% of offenders. In contrast, 95% of total crime was caused by 91% of offenders.

Yet, *counts* of crime are concentrated in *places* to a far greater degree than across suspected *offenders*. Considering the top 1% of streets and the top 1% of offenders, crime over 3 years was concentrated in the former at a rate of almost two and a half times than the latter.

The *harm* of crime is just the opposite. CHI harm is concentrated substantially more among a power few offenders than among a power few places. While the concentration of harm when considering the top 1% of streets and suspected offenders is broadly comparable (at 27.8% and 25%, respectively), the concentration of harm around suspected offenders compared to streets is far greater when considering the top 5% in each category. The top 5% of streets suffered 51.4% of total harm; the top 5% of offenders caused 66% of harm.

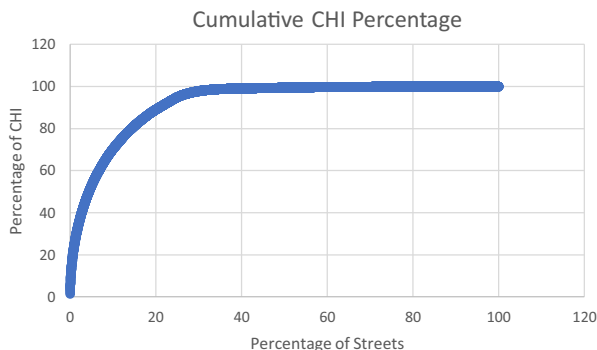


Fig. 9 Concentration of crime harm (Cambridge CHI) across all Cambridgeshire

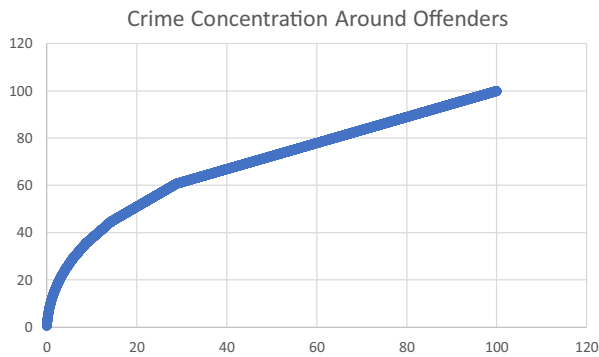


Fig. 10 Crime count concentration across confirmed suspects, all Cambridgeshire

Professional Judgment in Identifying High Crime and Harm Streets

Once it became clear that systematic analysis could identify the most prolific and harmful streets and offenders, the analysis proceeded to measure how well front line officers can identify these power few based on their experiential knowledge and professional judgement. Based on the series of workshops and interviews with 123 officers described above, the analysis compared the respondents' lists of perceived top ten 'hottest' streets for counts of crime *in their specific police area* to lists generated by the statistical analyses summarised in Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 above, identified the actual hottest streets, producing a mark out of ten.

For police officers across Cambridgeshire, the mean mark for correctly identifying the ten hottest streets for crime counts in their local area ($N = 123$) was 23%, while the PCSOs averaged 30%. The range of scores was between 0 and 60% correct. The results for identifying the ten streets with highest crime harm were similar at 26% correct, also with a range from 0 to 60%. The 14 PCSOs scored 35% correct.

For the identification of the most prolific of suspected offenders, the intuitions of 40 officers were gained in small group workshops in line with the methodology described above. The results were then compared to the top ten prolific suspected offenders for the area relevant to the individual workshop participant and the participant responses marked out of ten. The results were less accurate than for places, with a mean score for police officers across Cambridgeshire was 9%, and a range of scores between 0 and 30%.

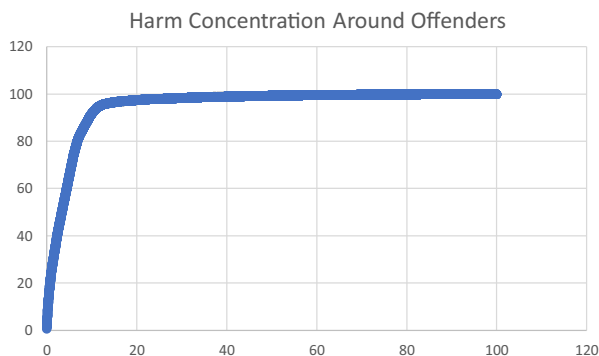


Fig. 11 Harm concentration across confirmed suspects, all Cambridgeshire

In naming the top ten most harmful suspected offenders, the 40 officers did about the same as in naming the most prolific. The mean concordance between the names they offered and the names identified by comprehensive data analysis was 5%, with a range between 0 and 10%.

Further analyses were then conducted of the sensitivity of these conclusions to length of officer experience, both in total police service and in time spent in the local area for which their judgement was requested (Sutherland 2017). These tests, using scatterplots for accuracy of listings of highest crime count streets and length of service, failed to find much correlation between the two variables for either total length of service (Pearson's $r = 0.049$, $p = 0.68$) or length of time working in the local area ($r = -0.099$; $p = 0.41$). Results for crime harm spots were similar.

Predictors of Accuracy

The study examined two potential predictors of how accurately the officers could identify power few targets based on professional judgement. One was self-confidence in the ability to do so accurately; the results showed that this variable had no predictive value of accurate identification. The other predictor was whether the officers worked in urban or semi-rural areas. The results showed that officers working in urban areas made more accurate identifications of high crime and harm streets than officers working in semi-rural areas.

During the workshops, participants were asked to assess their own confidence in being able to identify hotspots of crime and harm and (for police officers only) their confidence in being able to identify the most prolific and the most harmful suspected offenders. Pearson correlation coefficients were calculated to examine the relationship between participant confidence and intuitive accuracy in all four lists of streets and suspected offenders (Sutherland 2017: Table 10). None of the r values exceeded 0.17, and none were statistically significant (two-tailed test at .05).

In contrast, the analysis of differences in accuracy of professional judgement between urban and semi-rural officers found a large effect size (Cohen's $d = .9$, $p = 0.00$) for the first test, in which 42 urban officers averaged 27% accuracy of identifying high crime-count streets compared to only 17% accuracy for 30 semi-rural officers (Sutherland 2017). For identifying high *harm* streets, the urban officers did even better: 35% accuracy for 42 urban officers compared to 16% for 30 semi-rural officers, a two-tailed significant difference ($p = .000$) with Cohen's $d = 1.3$ (Sutherland 2017).

Discussion

These findings raise an important question: how accurate should we expect police officers' professional judgement to be compared to an evidence-based targeting analysis? There is no easy answer for this question, nor is there any professional benchmark. Yet, many people who challenge the need for evidence-based targeting analysis might find these results surprising, if not disappointing. If officers are targeting streets that need patrol the most with only 23% to 26% accuracy, that may mean that most of the patrol they provide is of far less value than it could be. Given the historical expectations that police officers should know what is happening on their beats, and

especially *where* it is happening, the results suggest that data analysis is needed to meet those expectations in the modern world. In a world of patrol almost exclusively by automobile, can it be reasonable to expect police officers to generate such accurate intuitions within the large areas they are expected to patrol?

Reasonable or not, this research suggests that they are not able to do so. This finding is consistent with Kahneman's (2011) analysis of system I vs system II modes of thinking. The authors have no doubt of the professional expertise of the research participants in many dimensions of policing, as Kahneman would suggest they can acquire. Yet, the ability to subconsciously manage and interpret big-data patterns to form accurate intuitions on crime patterns is not susceptible to such expertise, since the patterns are not stable across either officers or even places over multi-year periods (Weinborn 2017). This study has provided further support for the existing research that suggests severe limitations in officers' ability to identify hotspots based on their own experience-driven professional judgments, most notably supporting the findings in Macbeth (2015).

This study has gone even further than previous research, however, by examining officers' perceptions of the most prolific and harmful suspected offenders. Compared to a comprehensive, evidence-based analysis, the findings show that officers' identification by professional judgement of the most prolific suspected offenders are 91% incorrect. For the identification of the most harmful suspected offenders, their judgement was 95% incorrect. Given public interest in the prevention of crime by prolific and harmful offenders, it is striking that officers are even less accurate in identifying those offenders in their areas than they are in identifying high-crime streets. This was particularly the case for the most harmful suspected offenders: only two officers (out of 40) were able to identify a single suspected offender within the top ten most harmful in their areas.

The reason for these low scores is unclear, and not discernible by the present research methodology. In attempting to understand how officers arrived at their professional judgements, a frequency analysis was conducted on those streets most frequently identified by officers as being within the top ten for crime and harm but was in fact outside of the 'power few.' One interesting example of a frequently named street was in Peterborough City.

"Crabtree" is a collection of cul-de-sacs sprouting off Peterborough's main street, all under the same road name and comprising a small residential community. Why did this street feature heavily in officer's intuitions of both crime and harm? A review of the crime in Crabtree certainly reveals that it does suffer a high frequency of crime: 186 crime reports in the 3-year period under review. This allowed it to be ranked 48th in terms of its CHI score (12,810 days of imprisonment recommended as the CHI score) and 44th for crime counts. The crimes committed within this street were extraordinarily varied: frequent criminal damages to vehicles, burglaries, assaults, serious assaults, thefts of and from motor vehicles, and less common offences such as 'exposure' or sending malicious letters. However, it is not clear why Crabtree featured so heavily in officer's minds over streets with considerably more crime that featured less frequently in their answers. To answer this question, further research would have to be carried out on the nature of the 'power-few' streets and the types of crime committed there. One possible explanation is that officers are more likely to be dispatched and/or remember individual victims in a residential setting than streets with high volume of crime that feature business or night-time economy victims of crime.

It could be argued that expecting officers to be able to recall the names of offenders (as opposed to places, faces or crimes) on the spot and without aid is an unreasonable

task. Subsequent to the initial round of workshops, the author ran a one-off group workshop, this time bringing together a team of six detectives. They were set the same task as the frontline officers; however, this time, they were encouraged to work as a group, pool their collective knowledge and produce a group consensus. The results were not encouraging. The detectives failed to identify any of the most prolific suspected offenders in the top ten for their area and only 1 out of the 10 most harmful suspected offenders. It is reasonable to conclude that police officers do not retain accurate knowledge about the most prolific and harmful suspected offenders. This is operationally important if officers are either proactively self-tasking or creating local offender-based priorities without the aid of analytical methods. In such circumstances, the opportunity for accurate targeting is being lost.

Conclusions and Policy Implications

These findings have a number of policy implications. Firstly, the discretion of front line officers to patrol proactively based on professional judgement can become better informed by evidence-based analytical methods: officers cannot be reasonably expected to correctly identify hotspots of crime and harm without supporting data analysis. This information also needs to be given greater context and meaning by education and training for officers, so that they understand the power of analytical methods to help them with their work. Properly conceived, analytical methods can be explained as a tool to help officers, not take away their discretion. Failure to properly recognise the cultural values of discretion and autonomy seems likely to lead to rank and file rejection of analytical methods.

Police leaders must therefore find a way of combining an evidence-based approach with hard-earned officer experience. One approach to achieving this is to promote a recognition that analytical approaches will tell you *where the hotspots are, but not what to do* when you get there. How to effectively provide policing once inside a hotspot is a more appropriate subject for applying the experience and craft of frontline officers.

Secondly, police commanders from rural and semi-rural areas must be aware of the potential of a hot-spots approach to identify hitherto unknown concentrations of crime in their area. The evidence shows that these concentrations will be present, but that rural officers are less likely to be aware of them than urban officers without being supplied with the comprehensive evidence.

Thirdly, briefing and tasking systems must be designed on the premise that officers are unlikely aware of the most prolific and harmful suspected *offenders* in a given area. Less must be assumed, and more analytical products should be provided, in order to help officers identify the most prolific offenders. Such information (with photos) can be incorporated into shift briefings, in addition to other targets and priorities. The provisions must be combined with realistic expectations of what officers are supposed to do with this information. One possible bridge would include the application or discussion of other evidence-based practices that are proven to reduce the harm that offenders cause (Sherman 2013).

Fourthly, the existence of persistently hot streets should provoke further analysis as to the preventable causes of crime in that area. An analytical, problem-oriented policing approach can help inform and better target resources, even in rural areas.

Finally, police forces should undertake more experiments in the potential of technology to address these policy implications. Mechanical tracking of officers with GPS reports

seems likely to run afoul of policing culture. It may be preferable to use randomised trials or other tests of GPS systems automatically ‘nudging’ officers towards streets with the highest crime levels through the use of ‘push notifications’ and ‘gamification’ of rewards and feedback—much like Google traffic information does for all drivers. Nudging officers towards the hottest streets seems more likely to achieve the desired cultural change than a top-down authoritarian style. This is especially true given that the act of preventing crime by police presence does not provide the immediate feedback necessary to reinforce desired behaviours (Thaler and Sunstein 2009). Immediate feedback from a nudging computer screen may bring far more substantial change in patrol patterns.

The limitations of this study are presented in depth in Sutherland (2017). They are not so great, however, as to challenge the main conclusion and policy implication. Professional judgement can be enhanced, but not replaced, by the addition of evidence-based targeting of the most high-crime places and offenders. Whether these findings would be replicated in more metropolitan conditions, such as London, requires a replication of the present study. Until such research is done, however, there seems little basis to claim that evidence-based targeting is unnecessary for any officers in the UK, if not elsewhere.

The surprisingly low levels of accuracy in target identification presented in this study may themselves act as a spur to officers. It may whet their appetites to explore and accept a more evidence-based approach to their work. By recognising both the value and the limitations of their policing experience, they can promote more widespread provision of better tools to help them do their jobs. If evidence-based policing is a tool, then the results of this study may persuade more officers to pick it up and try it.

Acknowledgments The authors would like to thank the College of Policing, as well as former Chief Constable Alec Wood and the Chief Officer Group of Cambridgeshire Constabulary for their support of the research on which this article is based, which was led by the first author as a thesis submitted to the University of Cambridge in partial completion of the Master of Studies in Applied Criminology and Police Management at the Police Executive Programme, Institute of Criminology. They also thank all of the front line officers who took part in the research.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2012). An ex post facto evaluation framework for place-based police interventions. *Evaluation Review*, 35(6), 592–626.
- De Becker, G. (2000). *The gift of fear*. New York: Bloomsbury.
- Dudfield, G., Angel, C., Sherman, L. W., & Torrence, S. (2017). The “power curve” of victim harm: targeting the distribution of crime harm index values across all victims and repeat victims over 1 year. *Cambridge Journal of Evidence-Based Policing*, 1(1), 38–58.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.
- Hagan, F. (2006). *Research methods in criminal justice and criminology* (7th ed.). Boston: Allyn and Bacon.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
- Liggins, A. (2017). *Tracking the Most serious offenders in Northamptonshire: Continuity and replacement over time in the power few*. M.St. Thesis in applied criminology and police management, Institute of Criminology, University of Cambridge.

- Macbeth, E. (2015) *Evidence-based vs experience-based targeting of crime and harm hotspots in Northern Ireland*. (M.St. Thesis in applied criminology and police management, Institute of Criminology, University of Cambridge).
- Macbeth, E., & Ariel, B. (2017). Place-based statistical versus clinical predictions of crime hot spots and harm locations in Northern Ireland. *Justice Quarterly*, 1–34.
- Martin, S. E., & Sherman, L. W. (1986). Selective apprehension: a police strategy for repeat offenders. *Criminology*, 24(1), 155–174.
- Ratcliffe, J., & McCullagh, M. (2001). Crime, repeat victimisations and GIS. In Hirschfield, A., Bowers, K. (eds.) *Mapping and Analysing Crime Data* (p.61–92). London, UK: Taylor & Francis.
- Rengert, G. F., & Pelfrey, W. V. Jr. (1997). Cognitive mapping of the city center: Comparative perceptions of dangerous places. In Weisburd, D., & McEwen, T. (eds.) *Crime Mapping and Crime Prevention*, (p. 193–217). Monsey: Criminal Justice Press
- Sherman, L. (2007). The power few: Experimental criminology and the reduction of harm. *Journal of Experimental Criminology*, 3(4), 299–321.
- Sherman, L. (2013). The rise of evidence- based policing: Targeting, testing and tracking. In M. Tonry (Ed.), *Crime and Justice in America, 1975–2025, crime and justice: A review of research* (Vol. 42, pp. 377–452). Chicago: University of Chicago Press.
- Sherman, L., Gartin, P., & Michael, B. (1989). Hot spots of predatory crime: routine activities and the criminology of place. *Criminology*, 27(1), 27–55.
- Sherman, L., Neyroud, P., & Neyroud, E. (2016). The Cambridge crime harm index: Measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice*, 10(3), 171–183.
- Sutherland, J. (2017). Evidence vs. Intuition in identifying “power few” crime targets: A comparative analysis in multiple Neighbourhood policing areas. M.St. Thesis in applied criminology and police management, Institute of Criminology, University of Cambridge.
- Thaler, R., & Sunstein, C. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.
- Weinborn, C. (2017). From hotspots to harmspots, an empirical exploration of crime concentrations and crime harm in the United Kingdom. PhD dissertation, Institute of Criminology, University of Cambridge.
- Weinborn, C., et al. (2017). Hotspots vs harmspots: shifting the focus from counts to harm in the criminology of place. *Applied Geography*, 86, 226–244.
- Weisburd, D. (2015). The law of concentrations and the criminology of space. *Criminology*, 53(2), 133–157.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

James Sutherland is a serving Superintendent in Cambridgeshire Constabulary. He completed this research as part fulfilment of an M.St Degree in Applied Criminology and Police Management at Cambridge University and also holds a BA (Hons) in Politics from the University of York, an M:Litt in Strategic Studies from the University of Aberdeen and an MA in Management from the University of York. James is a member of the International Society of Crime Prevention Practitioners.

Affiliations

James Sutherland¹ · Katrin Mueller-Johnson²

✉ Katrin Mueller-Johnson
kum20@cam.ac.uk

¹ Cambridgeshire Constabulary, Cambridge, UK

² Institute of Criminology, University of Cambridge, Cambridge, UK